

Genetic variation of *Cytochrome P450 1B1 (CYP1B1)* and risk of breast cancer among Polish women

Mia M. Gaudet^a, Stephen Chanock^{a,b}, Jolanta Lissowska^{a,c}, Sonja I. Berndt^a, Xiaohong (Rose) Yang^a, Beata Peplonska^d, Louise A. Brinton^a, Robert Welch^b, Meredith Yeager^b, Alicja Bardin-Mikolajczak^c, Mark E. Sherman^a, Thomas R. Sutter^{e,f} and Montserrat Garcia-Closas^a

Four single nucleotide polymorphisms (SNPs) in *CYP1B1* (Ex2 + 143 C>G, Ex2 + 356 G>T, Ex3 + 251 G>C, Ex3 + 315 A>G) cause amino acid changes (R48G, A119S, L432V and N453S, respectively) and are associated with increased formation of catechol estrogens; however, epidemiologic evidence only weakly supports an association between these variants and breast cancer risk. Because genetic variability conferring increased susceptibility could exist beyond these putative functional variants, we comprehensively examined the common genetic variability within *CYP1B1*. A total of eight haplotype-tagging (ht)SNPs (including Ex3 + 315 A>G), in addition to two putatively functional SNPs (Ex2 + 143 C>G and Ex3 + 251 G>C), were selected and genotyped in a large case-control study of Polish women (1995 cases and 2296 controls). Haplotypes were estimated using the expectation-maximization algorithm, and overall differences in the haplotype distribution between cases and controls were assessed using a global score test. We also evaluated levels of tumor *CYP1B1* protein expression in a subset of 841 cases by immunohistochemistry, and their association with genetic variants. In the Polish population, we observed two linkage disequilibrium (LD)-defined blocks. Neither haplotypes (global *P*-value of 0.99 and 0.67 for each block of LD, respectively), nor individual SNPs (including three putatively functional SNPs) were associated with breast cancer risk. *CYP1B1*

was expressed in most tumor tissues (98%), and the level of expression was not related to the studied genetic variants. We found little evidence for modification of the estimated effect of haplotypes or individual SNPs by age, family history of breast cancer, or tumor hormone receptor status. The present study provides strong evidence against the existence of a substantial overall association between common genetic variation in *CYP1B1* and breast cancer risk. *Pharmacogenetics and Genomics* 16:547–553 © 2006 Lippincott Williams & Wilkins.

Pharmacogenetics and Genomics 2006, 16:547–553

Keywords: breast cancer, *CYP1B1*, family history, haplotypes, P450s

^aDivision of Cancer Epidemiology and Genetics, National Cancer Institute, Department of Health and Human Services, Bethesda, Maryland, USA, ^bCore Genotype Facility at the Advanced Technology Centre, National Cancer Institute, Department of Health and Human Services, Bethesda, Maryland, USA, ^cDepartment of Cancer Epidemiology and Prevention, Cancer Center and M. Skłodowska-Curie Institute of Oncology, Warsaw, Poland, ^dDepartment of Occupational and Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland, ^eDepartment of Environmental Health Sciences, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland, USA and ^fW. Harry Feinstone Centre for Genomic Research, University of Memphis, Memphis, TN, USA.

Correspondence and requests for reprints to Mia M. Gaudet, National Cancer Institute, 6120 Executive Blvd, EPS/7055, Rockville, MD 20852, USA. Tel: +1 301 435 4725; fax: +1 301 402 0916; e-mail: gaudetm@mail.nih.gov

Received 16 December 2005 Accepted 16 February 2006

Introduction

The initial step of estrogen metabolism is the conversion of estradiol (E₂) to 2-hydroxy-E₂ (2-OHE₂) and 4-hydroxy-E₂ (4-OHE₂) by the cytochrome P450 family of enzymes, of which *CYP1B1* is the most efficient member [1]. The 4-OHE₂ catechol metabolite has the highest carcinogenic as well as estrogenic activity and is found in higher quantities in breast cancer tissue than other E₂ metabolites, including 2-OHE₂ and C-16 α hydroxylation [1]. OHE₂ metabolites can be further converted to quinones and semiquinones, which cycle in a redox reaction generating reactive oxygen species. Over-expression of *CYP1B1* could lead to an accumulation of estrogen metabolites that increase cell proliferation and/or act directly as carcinogens [2]. Interindividual variability of estrogen metabolism could thus contribute to the risk of breast cancer [3].

CYP1B1 is located on chromosome 2p21–22 and has four common single nucleotide polymorphisms (SNPs) that encode amino acid substitutions (at codons 48, 119, 432 and 453) that increase catalytic activity [4,5]. Several studies examined these functional variants in relation to breast cancer risk [6]. In a recent meta-analysis of 5712 cases and 5107 controls, no association with these putative functional variants was found [6]. However, some studies have observed evidence that *CYP1B1* variants may be related to tumor estrogen receptor status [6–8].

Previous studies of *CYP1B1* and breast cancer have not considered the possibility that other variants in the gene, such as those in regulatory regions, or that particular haplotypes could alter risk. Therefore, to comprehensively assess common genetic variation within *CYP1B1*,

we assayed haplotype-tagging (ht)SNPs, including Ex3 + 315 A > G (N453S), and two additional SNPs, Ex2 + 143 C > G (R48G) and Ex3 + 251 G > C (L432V), with possible functional significance, and examined them in relation to breast cancer risk in a case-control study of Polish women. Haplotypes, inferred from the genotyped SNPs, and individual SNPs were evaluated, and risk estimates were stratified by age, family history of breast cancer and hormone receptor status. In addition, we determined CYP1B1 protein expression levels in breast tumor tissue samples of breast cancer cases, and evaluated whether level of expression is related to genetic variants.

Methods

Study population

A large population-based case-control study was conducted among women residing in two Polish cities, Warsaw and Lodz [9]. Institutional Review Board approval was obtained from all participating institutions, and signed informed consent was obtained for all respondents. Eligible cases were women aged 20–74 years who were newly diagnosed with either histologically or cytologically confirmed *in situ* or invasive breast cancer. Study personnel identified cases from January 2000 to January 2003 through a rapid identification system at participating hospitals that covered approximately 90% of all eligible cases. Controls with no history of breast cancer were randomly selected from January 2000 till September 2003 through the Polish Electronic System, a database of all Polish residents. Controls were frequency matched to cases by city and age in 5-year categories.

A total of 2386 cases (79% of the 3037 eligible cases identified) and 2502 controls (69% of the 3639 eligible controls identified) women provided a personal interview on known and suspected risk factors. The primary reasons for non-participation for cases and controls, respectively, were refusal (18% and 24%) and inability to locate the individual (2%, 65). Subjects refused to participate because of emotional distress (14% and 0%), lack of interest (16%, 26%), had no time (7%, 17%), or had other (4%, 3%) or unknown (60%, 54%) reasons. Trained nurses also collected venous blood samples from 1995 cases (84% of participating cases) and 2296 controls (94% of participating controls). Genomic DNA was isolated from buffy coats by the Autopure LS DNA Purification System (Gentra Systems, Inc., Minneapolis, Minnesota, USA).

Genotyping

The choice of htSNPs for *CYP1B1* was based on a strategy developed by the Breast and Prostate Cohort Consortium [10]. SNPs were identified through two parallel approaches: (i) SNPs were identified through the public database, dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) and (ii) to identify unknown missense SNPs, a resequence analysis of exonic regions of *CYP1B1* was

conducted in 190 patients with advanced breast or prostate cancer in the Multi-Ethnic Cohort [10]. Twenty-six SNPs identified and selected through these two approaches had a reported minor allele frequency > 5% in European-Americans and were spaced approximately every 1–2 kb. These SNPs were then genotyped in a multi-ethnic panel of 733 individuals (the BPC3 haplotyping panel) [10] to assess allele frequencies and impute haplotypes. Each assay was validated by resequencing analysis in the SNP500 cancer set of 102 individuals. Optimized assays were developed for 14 SNPs (rs163077, rs163086, rs162549, rs9341266, rs10916, rs162562, rs1800440, rs1056837, rs1056836, rs162560, rs162557, rs162556, rs162555 and rs10175368). An additional 12 possible SNPs could not be reliably assayed on any of three genotype platforms available (TaqMan, EPOCH or Sequenom) or were determined to have MAFs well below 5%. Common haplotypes were determined among the 206 European-Americans in the BPC3 haplotyping panel [including 172 unrelated parent Centre d'Etude du Polymorphisme Humain (CEPH) samples] using Excoffier's and Slatkin's expectation-maximization algorithm [11], and haplotype structure was confirmed among 12 three-generation CEPH family pedigrees. Two linkage disequilibrium (LD)-defined blocks in *CYP1B1* were identified. The TagSNPs program (<http://www-rcf.usc.edu/~stram/>) was used to select a minimum set of htSNPs with a pairwise $r^2 > 0.8$ [12]. Due to the extensive linkage disequilibrium across *CYP1B1*, eight htSNPs (Table 1) were chosen to capture the majority of common haplotype diversity among Caucasians ($r^2_h = 0.996$).

The approach for determining htSNPs included SNPs with a MAF > 5% but did not include all non-synonymous SNPs and in particular, those studied in breast cancer [6]. Of the four non-synonymous SNPs with possible functional consequences (Ex2 + 143 C > G, R48G; Ex2 + 356 G > T, A119A; Ex3 + 251 G > C, L432V; Ex3 + 315 A > G, N453), one at Ex3 + 315 A > G was selected as an htSNP. For completeness and comparison with previous studies, we also genotyped Ex2 + 143 C > G and Ex3 + 251 G > C, but did not analyse the SNP at Ex2 + 356 G > T because neighbouring SNPs hindered assay optimization. However, strong LD across the gene suggests that Ex2 + 356 G > T was analysed indirectly, especially because other studies have reported that Ex2 + 143 C > G and Ex2 + 356 G > T are in strong LD [13,14].

Genotype analysis was performed in 1995 cases and 2296 controls. A total of 100 duplicate DNA pairs were interspersed throughout the DNA samples. All pairs were greater than 98% concordant for each SNP with the exception of Ex2 + 143 C > G, for which 93% of the pairs were concordant. The DNA plates also contained 41 replicate samples for two unidentified donors. These

Table 1 Description of *CYP1B1* haplotype-tagging (ht) and additional non-synonymous single nucleotide polymorphisms (SNPs)

No.	Location	Nucleotide substitution	Minor allele frequency	Landmark	rs#	Block # for htSNP
1	IVS10-8520 ^a	T>C	0.26		rs163077	1
2	IVS10-1363 ^a	T>C	0.20		rs163086	1
3	Ex3-1249	C>T	0.03	3'-UTR	rs9341266	1
4	Ex3 + 939	A>C	0.22	3'-UTR	rs162562	1
5	Ex3 + 315	A>G	0.16	N453S	rs1800440	1
6	-3922	C>T	0.44		rs162556	2
7	-5329	G>A	0.33		rs10175368	2
8	-2919	C>T	0.25		rs162557	2
—	Ex3 + 143	C>G	0.33	R48G	rs10012	N/A
—	Ex3 + 251	G>C	0.42	L432V	rs1056836	N/A

^aLocated within gene FLJ32954.

replicates were 100% concordant with the exception of Ex2 + 143 (concordance was 91%). Description and methods for each genotype assay can be found at <http://snp500cancer.nci.nih.gov> [15]. Genotype frequencies for all loci were in Hardy–Weinberg equilibrium among controls ($P > 0.18$).

Immunostaining

Tissue microarrays (TMA) were constructed from paraffin-embedded formalin-fixed tumor blocks from 841 cases. Two separate 0.6-mm tissue cores for each case were extracted and placed on two TMAs. TMAs were then stained for CYP1B1 expression using the rabbit-derived anti-P4501B1 IgG, as previously described [16]. A single pathologist rated the CYP1B1 expression using a 0–3 intensity score, for 766 cases with informative cores. We created a dichotomized variable from the intensity scores of each core (scores 0 and 1 versus scores 2 and 3; the agreement between cores was 77%), and each subject was assigned the maximum level of expression for the two cores. The associations between the expression score and genetic variants were tested using Pearson's chi-square.

Statistical analysis

Pairwise LD was estimated between htSNPs based on D' values [17,18] using Haploview (<http://www.broad.mit.edu/mpg/haploview/index.php>). Block structure for the purpose of statistical analyses was determined using the genotype data from the Polish controls and defined based on the solid spine algorithm using $D' > 0.80$ as the threshold cut-off; however, the block definition was robust to changes in the D' cut-off and inclusion of the putatively functional SNPs, Ex2 + 143 C > G and Ex3 + 251 G > C (data not shown). We also examined alternative block structures, including combining all htSNPs in one block and the block structure used to select htSNPs (detailed above). To assess possible *cis*-effects, we examined haplotypes that included the three putatively functional SNPs.

For each block, haplotype frequencies and associated measures of effect were estimated using HaploStats (version 1.2.1; <http://mayoresearch.mayo.edu/mayo/>

[research/schaid_lab/](http://research.schaid_lab/)), which employs the expectation-maximization (EM)-algorithm to estimate haplotype frequencies and an iterative two-step EM model to estimate the association between individual haplotypes and risk assuming an additive model [19,20]. Models were run separately by haplotype block. A global score statistic, adjusted for the matching factors age (in 5-year categories) and study site (Lodz or Warsaw), was used to evaluate the overall difference in haplotype frequencies between cases and controls [20].

Unconditional logistic regression models were used to estimate odd ratios (OR) and 95% confidence intervals (CI), adjusted for age and study site, for the association between individual *CYP1B1* polymorphisms and breast cancer, using STATA (version 8.2; STATACorp, College Station, Texas, USA). We evaluated the data using indicator variables for the genotypes, and assessed the fit of an additive (analogous to the P -value for linear trend) genetic model using a single variable coded for the number of variant alleles present.

Individual SNP and haplotype results were stratified by age (50 years or less, over 50 years) and history of a first-degree relative diagnosed with breast cancer (yes/no) because younger women as well as women with a family history of breast cancer could be at increased risk for genetic risk factors. We also analysed whether the estimated association was modified by menopausal status and body mass index (BMI) (by categories of menopausal status). To assess deviations from a multiplicative interaction model, we used the log-likelihood ratio test to compare the fit of logistic models with and without an interaction term.

Breast cancer cases were further defined by estrogen (ER+, ER-) and progesterone (PR+, PR-) tumor receptor status. Hormone receptor status was determined by immunohistochemical assay for 91% of cases. Among women with *CYP1B1* genotype data, a total of 910 (66%) of cases were diagnosed with an ER+ tumor, and 760 (55.5%) cases were diagnosed with a PR+ tumor. Cases with unknown hormone status were dropped from these analyses. Polytomous regression models were used to

simultaneously estimate ORs and 95% CI for two different case groups compared to controls. In addition, we tested for heterogeneity of the genotype OR using a case only design in which the different tumor subtypes were the outcome variable for the logistic regression models.

We examined factors that were related to breast cancer risk in this study population as potential confounders, including education, age at menarche, parity, type of menopause, age at first full-term pregnancy, age at menopause, BMI, family history of breast cancer, and personal history of benign breast disease. We employed a backwards elimination approach and potential confounders did not alter estimates of the observed associations (for individual genotypes or haplotypes) by more than 10%; therefore, final models controlled for matching factors (age and study site) only.

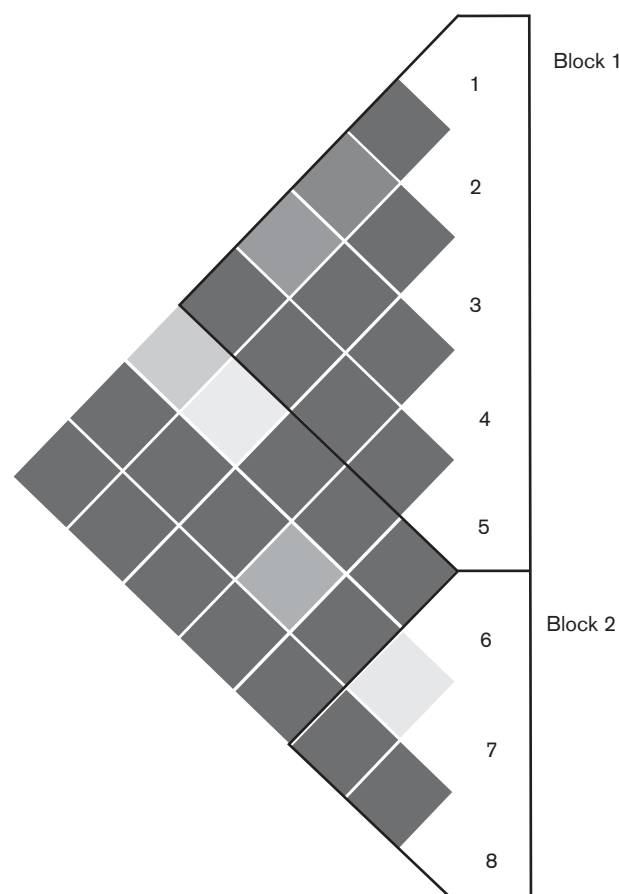
Results

In our case-control study of breast cancer in Poland, we observed that the established risk factors (e.g. age at menarche, age at and number of full-term births, history of benign breast disease, and family history of breast cancer in first degree relatives) were associated with breast cancer risk in comparable direction with similar estimates of magnitude reported in previous studies [9]. The cases were diagnosed at a mean \pm SD age of 56 ± 10 years with stage I (12.2%), stage II (60.8%), or stage III and above (27.0%). Examination of LD blocks (Fig. 1) among this study population revealed two blocks. These blocks were tagged by eight SNPs, including five SNPs for block 1 and three SNPs for block 2. Of the 32 possible haplotype combinations for block 1, five common haplotypes were observed. For block 2, five common haplotypes were observed in our population.

None of the *CYP1B1* haplotypes were significantly associated with breast cancer risk compared to women carrying the most common haplotype within each block (Table 2). Global tests within each block also indicated a lack of an association (Table 2). The results obtained did not alter when we examined alternative haplotype block definitions (data not shown). There was no evidence of *cis*-effects between the putatively functional SNPs (data not shown).

Haplotype associations with breast cancer did not vary by age, menopausal status, BMI, ER status or PR status (results not shown). However, when we stratified results by family history of breast cancer, we observed a decreased risk of breast cancer associated with haplotype 1b (OR = 0.63, 95% CI 0.28–1.03), which contains the variant allele for the Ex3 + 315 A > G polymorphism, among women with a family history of breast cancer. This haplotype was not associated with risk among women

Fig. 1



The linkage disequilibrium (LD) plot of *CYP1B1* for a case-control study of Polish women. The darkest shade represents a D' of 0.9 or higher. The remaining lighter shades represent lesser degrees of D' . The numbering of each locus corresponds to the numbering of SNPs in Table 1.

without a family history of breast cancer (OR = 1.05, 95% 0.91–1.20). The Ex3 + 315 G variant allele was also carried on haplotype 1h; however, this haplotype was not present among women with a family history of breast cancer. When we examined results stratified by family history for the Ex3 + 315 GG variant genotype, the results were consistent with the results for haplotype 1b.

None of the putatively functional SNPs (Table 3) was associated with breast cancer, and associations were similar across stratum by age, family history, menopausal status, and BMI of breast cancer for all SNPs, except Ex3 + 315 A > G (results not shown). Consistent with the results for haplotype 1b, when we examined results for the Ex3 + 315 A > G polymorphism stratified by family history, the variant G allele was associated with a decreased risk of breast cancer among women with a family history of breast cancer (OR = 0.60, 95% CI 0.37–0.98). Examination of SNPs by estrogen and progesterone

Table 2 Age- and centre-adjusted odds ratios (OR) and 95% confidence intervals (CI) for the association between estimated *CYP1B1* haplotypes and breast cancer among a case-control study of Polish women (1995 cases and 2296 controls)

CYP1B1 htSNP haplotypes						Haplotype frequency		Haplotype-specific ^a	
	IVS10-8520 T>C	IVS10-1363 T>C	Ex3-1249 C>T	Ex3+939 A>C	Ex3+315 A>G	Cases	Controls	OR	(95% CI)
Block 1									
1a	C	C	C	A	A	31.1	31.1	1.00	
1b	C	C	C	A	G	16.1	16.1	1.01	(0.88–1.15)
1c	C	C	C	C	A	4.1	4.1	1.00	(0.78–1.27)
1d	C	C	T	A	A	3.0	2.9	1.05	(0.81–1.37)
1e	C	T	C	A	A	20.5	19.9	1.03	(0.91–1.16)
1f	T	C	C	A	A	7.6	8.1	0.93	(0.78–1.12)
1g	T	C	C	C	A	17.6	17.6	1.00	(0.88–1.14)
1h	T	C	C	A	G	0.0	0.4	0.86	(0.52–1.44)
1i	T	C	T	A	A	0.5	0.7		
Global <i>P</i> -value						0.99			
Block 2									
2a	C	C	A			32.7	32.5	1.00	
2b	C	C	G			17.7	16.9	1.04	(0.91–1.19)
2c	C	T	G			25.0	25.8	0.96	(0.86–1.08)
2d	T	C	G			6.2	6.7	0.91	(0.74–1.13)
2e	T	T	G			18.4	18.1	1.01	(0.90–1.16)
2f	C	T	A			0.0	0.1	N/E	
2g	T	T	A			0.0	0.0		
Global <i>P</i> -value						0.67			

^aEstimates are based on an additive effect model.**Table 3** Age- and centre-adjusted odds ratios (OR) and 95% confidence intervals (CI) for the association between *CYP1B1* non-synonymous single nucleotide polymorphisms (SNPs) and breast cancer among a case-control study of Polish women (1995 cases and 2296 controls)

SNP location (amino acid change)	Genotype	Cases		Controls		OR	(95% CI)	<i>P</i> -value	<i>P</i> -value for trend
		<i>n</i>	%	<i>n</i>	%				
Ex2+143C>G (R48G)	CC	910	46.4	1057	46.8	1.00			
	CG	805	41.0	921	40.8	1.02	(0.89–1.16)	0.79	
	GG	248	12.6	280	12.4	1.03	(0.85–1.25)	0.78	0.74
Ex3+251G>C (L432V)	CC	652	34.0	772	34.7	1.00			
	CG	913	47.6	1033	46.4	1.05	(0.91–1.20)	0.50	
	GG	353	18.4	419	18.8	0.99	(0.83–1.18)	0.92	0.95
Ex3+315A>G (N453S)	AA	1338	67.1	1538	67.0	1.00			
	AG	504	25.3	573	25.0	1.01	(0.88–1.17)	0.85	
	GG	55	2.8	64	2.8	0.98	(0.68–1.42)	0.93	0.92

receptor status did not reveal any appreciable heterogeneity of risk estimates (results not shown).

Of the 766 breast cancer cases with CYP1B1 protein expression scores, nearly all tumor tissue samples (98.1%) expressed the CYP1B1 protein with 24.5%, 53.8% and 19.8% of tumors with low, medium and high expression scores, respectively. We found no association between levels of expression and individual polymorphisms, nor haplotypes (data not shown).

Discussion

Previous studies have examined four common, putatively functional SNPs in *CYP1B1* [6], several of which have been associated with increased enzyme activity in some studies; however, the results from a recent meta-analysis

indicates these polymorphisms are probably not associated with breast cancer [6]. Because genetic variability conferring increased susceptibility may exist beyond these putative functional SNPs, we comprehensively examined the genetic variability within *CYP1B1* by selecting and genotyping htSNPs in a large case-control study in Poland. Our results show that common genetic variation within *CYP1B1* was not associated with overall breast cancer risk. In addition, the lack of association found in the present study for three of the putative SNPs at codons 48, 432, and 452 confirms the lack of association reported in a recent meta-analysis [6].

In initial reports, it was observed that amino acid substitutions at codons 48, 119, 432, and 453 were associated with 2.4- to 3.4-fold higher catalytic activity

compared to the wild-type enzyme [4,5]; however, additional recent research suggests rare mutations [21], rather than common polymorphisms, are more strongly associated with alterations in activities [21,22]. It is interesting to note that the regions in which the putative non-synonymous SNPs reside are not highly conserved in mammals with the exception of the SNP within codon 453 [21]. It is likely that variants residing in regulatory regions could alter the expression or stability of the *CYP1B1* gene [23]. For example, it is possible that SNPs in an estrogen response element proposed to reside between bps -63 and -49 of the promoter region, or xenobiotic response elements proposed to lie at -834 and -853 of the 5-flanking region could be of great interest [24,25]. Our strategy of htSNP selection did not identify SNPs with high MAFs in these regions, although it is possible that lower frequency variants could be functionally important, and perhaps are associated with cancer risk.

We examined possible effect modification by age, family history of breast cancer, and tumor hormone receptor status, because of *a priori* expectations that genetic susceptibility might be more evident among women with a family history of breast cancer, early onset breast cancer [26], or expressing ER and PR [27] in the tumor tissue. We found the variant Ex3 + 315 G allele (453S) was associated with a lower risk of breast cancer among women with a family history of breast cancer, but not among women without family history of breast cancer. This result is apparently consistent with a Swedish population-based case-control study (1500 cases, 1500 controls) [28]; however, the results of the present study were driven by differences in the distribution of the variant allele among cases by family history of breast cancer, whereas the results of the Swedish study [28] were driven by differences in the controls. In addition, these results are inconsistent with our *a priori* hypothesis that these variants would be associated with an increased risk of breast cancer. Some studies found that the frequency of individual SNPs was associated with ER status [7,8,28]; however, we found no heterogeneity in risk estimates when we examined ER and PR status.

The prevalence of CYP1B1 expression in this population-based study was consistent with previous studies of mRNA tissue expression in clinical samples [29]. Although we did not find an association between *CYP1B1* variants and protein expression, CYP1B1 was expressed in the tissue samples of most case women in our study. A comparison of the level of expression as measured by intensity of staining in two distinct tissue cores showed low concordance, which could be partly due to heterogeneity in the immunohistochemistry staining of the tumor tissue. However, when we dichotomized the intensity score, the agreement between the two cores improved to 76.5%.

The present study has adequate power to evaluate either an additive or dominant model for disease association. Based on the observed frequency estimates for either genotype analysis or haplotype analysis, we had at least 80% power to detect an OR of 1.3 or higher assuming an additive model. However, the power to detect a recessive genotype association was substantially lower, particularly for low frequency alleles such as Ex 3 + 315 (N453S); see the wide confidence intervals in Table 3. Our study also benefited from some of the highest participation rates attained in molecular epidemiologic studies with the collection of blood, although we cannot exclude the possibility of selection bias [30]. However, the potential bias would not be expected to substantially affect our results because carrier status is unlikely to be associated with reasons for non-participation. In addition, we observed that the frequency of alleles, as well as the magnitude of effect of well-defined breast cancer risk factors (data not shown), was consistent with previous studies.

Population stratification is a concern in some studies; however, it is of minimal concern for the results obtained in the present study because our study population consisted of ethnically homogenous Polish women. Our conclusions are restricted to an analysis of *CYP1B1* and not other genes in the estradiol metabolism pathway. It is feasible that altered enzyme capacity of CYP1B1, caused by polymorphisms, may interact with causal variants in other relevant genes (e.g. *CYP1A1*). However, because our estimates of the main effects were very close to unity, such subgroup associations must be of small magnitude, or 'at risk' subgroups must be uncommon, unless there are cross-over effects, which are unlikely to occur [31].

Our haplotype-tagging selection strategy allowed us to comprehensively assess common genetic variation within *CYP1B1*. Based on this strategy, we found that common genetic polymorphisms within this gene alone are unlikely to contribute to breast cancer risk. It is possible that rare variants in *CYP1B1* could be associated with breast cancer risk [32], but studies with larger numbers, will be needed to address this possibility. Evaluation of subtle associations between *CYP1B1* and breast cancer in a small subgroup of women, such as those with a family history of breast cancer, will also require a larger sample size to be confirmed.

Acknowledgements

This research was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA and a Department of Defence grant (DOD DAMD 17-03-1-0229). The authors would like to thank Neonila Szeszenia-Dabrowska of the Nofer Institute of Occupational Medicine (Lodz, Poland) and Witold Zatonski of the M. Sklodowska-Curie Institute of

Oncology and Cancer Center (Warsaw, Poland) for their contribution to the Polish Breast Cancer Study. Anita Soni (Westat, Rockville, Maryland, USA) and Pei Chao (IMS, Silver Spring, Maryland, USA) have been invaluable to the management of the study. This work would not be possible without the dedicated efforts of the physicians, nurses, interviewers and study participants.

References

- Hayes CL, Spink DC, Spink BC, Cao JQ, Walker NJ, Sutter TR. 17 beta-estradiol hydroxylation catalyzed by human cytochrome P450 1B1. *Proc Natl Acad Sci USA* 1996; **93**:9776–9781.
- Yager JD. Endogenous estrogens as carcinogens through metabolic activation. *J Natl Cancer Inst Monogr* 2000; **27**:67–73.
- Creveling CR, Kirk KL. The effect of ring-fluorination on the rate of O-methylation of dihydroxyphenylalanine (DOPA) by catechol-O-methyltransferase: significance in the development of 18F-PETT scanning agents. *Biochem Biophys Res Commun* 1985; **130**:1123–1131.
- Hanna IH, Dawling S, Roodi N, Guengerich FP, Parl FF. Cytochrome P450 1B1 (CYP1B1) pharmacogenetics: association of polymorphisms with functional differences in estrogen hydroxylation activity. *Cancer Res* 2000; **60**:3440–3444.
- Shimada T, Watanabe J, Kawajiri K, Sutter TR, Guengerich FP, Gillam EM, Inoue K. Catalytic properties of polymorphic human cytochrome P450 1B1 variants. *Carcinogenesis* 1999; **20**:1607–1613.
- Wen W, Cai Q, Shu XO, Cheng JR, Parl F, Pierce L, et al. Cytochrome P450 1B1 (CYP1B1) polymorphism with steroid receptor status in breast cancer study and a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2005; **14**:329–335.
- De Vivo I, Hankinson SE, Li L, Colditz GA, Hunter DJ. Association of CYP1B1 polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002; **11**:489–492.
- Bailey LR, Roodi N, Dupont WD, Parl FF. Association of cytochrome P450 1B1 (CYP1B1) polymorphism with steroid receptor status in breast cancer. *Cancer Res* 1998; **58**:5038–5041.
- Garcia-Closas M, Brinton LA, Lissowska J, Chatterjee N, Peplonska B, Anderson W, et al. Established breast cancer risk factors by clinically important tumor characteristics. 2006; submitted.
- The National Cancer Institute Breast and Prostate Cancer Cohort Consortium. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer* 2005; **5**:977–985.
- Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**:921–927.
- Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC. Choosing haplotype-tagging SNPs based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered* 2003; **55**:27–36.
- Stoilov I, Akarsu AN, Alozie I, Child A, Barsoum-Homsy M, Turacli ME, et al. Sequence analysis and homology modeling suggest that primary congenital glaucoma on 2p21 results from mutations disrupting either the hinge region or the conserved core structures of cytochrome P4501B1. *Am J Hum Genet* 1998; **62**:573–584.
- Murray GI, Taylor MC, McFadyen MC, McKay JA, Greenlee WF, Burke MD, Melvin WT. Tumor-specific expression of cytochrome P450 CYP1B1. *Cancer Res* 1997; **57**:3026–3031.
- Packer BR, Yeager M, Staats B, Welch R, Crenshaw A, Kiley M, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res* 2004; **32**:D528–D532.
- Kim JH, Sherman ME, Curriero FC, Guengerich FP, Strickland PT, Sutter TR. Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. *Toxicol Appl Pharmacol* 2004; **199**:210–219.
- Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; **3**:299–309.
- Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004; **5**:89–100.
- Schaid DJ, Rowland CM. Robust transmission regression models for linkage and association. *Genet Epidemiol* 2000; **19** (Suppl 1):S78–S84.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002; **70**:425–434.
- Mammen JS, Pittman GS, Li Y, Abou-Zahr F, Bejjani BA, Bell DA, et al. Single amino acid mutations, but not common polymorphisms, decrease the activity of CYP1B1 against (–)benzo[a]pyrene-7R-trans-7,8-dihydrodiol. *Carcinogenesis* 2003; **24**:1247–1255.
- Han W, Pentecost BT, Spivack SD. Functional evaluation of novel single nucleotide polymorphisms and haplotypes in the promoter regions of CYP1B1 and CYP1A1 genes. *Mol Carcinog* 2003; **37**:158–169.
- Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405**:847–856.
- Tsuchiya Y, Nakajima M, Yokoi T. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett* 2005; **227**:115–124.
- Tsuchiya Y, Nakajima M, Kyo S, Kanaya T, Inoue M, Yokoi T. Human CYP1B1 is regulated by estradiol via estrogen receptor. *Cancer Res* 2004; **64**:3119–3125.
- Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. *Nat Rev Cancer* 2004; **4**:850–860.
- Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiol Biomarkers Prev* 2004; **13**:1558–1568.
- Rylander-Rudqvist T, Wedren S, Granath F, Humphreys K, Ahlberg S, Weiderpass E, et al. Cytochrome P450 1B1 gene polymorphisms and postmenopausal breast cancer risk. *Carcinogenesis* 2003; **24**:1533–1539.
- Modugno F, Knoll C, Kanbour-Shakir A, Romkes M. A potential role for the estrogen-metabolizing cytochrome P450 enzymes in human breast carcinogenesis. *Breast Cancer Res Treat* 2003; **82**:191–197.
- Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006; **163**:197–203.
- Pharoah PD, Dunning AM, Ponder BA, Easton DF. The reliable identification of disease-gene associations. *Cancer Epidemiol Biomarkers Prev* 2005; **14**:1362.
- Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**:124–137.